

AF/2655 *zfw*



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

ELIOT M. CASE

Group Art Unit: 2655

Examiner: James S. Wozniak

Serial No.: 09/871,524

Filed: May 31, 2001

For: METHOD OF TRAINING A COMPUTER
SYSTEM VIA HUMAN VOICE INPUT

Attorney Docket No.: 1813 (USW 0611 PUS)

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents
Commissioner for Patents
U.S. Patent & Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is an Appeal Brief in support of the appeal from the final rejection of claims 1-3, 6-7, 11-13, and 16-17 in the Office Action mailed on July 25, 2005.

I. REAL PARTY IN INTEREST

The real party in interest is Qwest Communications International Inc. ("Assignee"), a corporation organized and existing under the laws of the state of Delaware, and having a place of business at 1801 California Street, 38th Floor, Denver, Colorado 80202.

CERTIFICATE OF MAILING UNDER 37 C.F.R. § 1.8 (FIRST CLASS MAIL)

I hereby certify that this paper, including all enclosures referred to herein, is being deposited with the United States Postal Service as first-class mail, postage pre-paid, in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, U.S. Patent & Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 on:

1/26/06
Date of Deposit

Jeremy J. Curcuri
Name of Person Signing

Jeremy J. Curcuri
Signature

II. RELATED APPEALS AND INTERFERENCES

There are no appeals or interferences known to the Appellant, the Appellant's legal representative, or the Assignee which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1-3, 6-7, 11-13, and 16-17 are pending in this application. Claims 1-3, 6-7, 11-13, and 16-17 have been rejected and are the subject of this appeal.

IV. STATUS OF AMENDMENTS

An response after final rejection was filed on September 15, 2005. The response did not include any amendments.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention relates to a method of training a computer system via human voice input from a human teacher, with the computer system including a speech recognition engine. *Specification, p. 1, ll. 6-8.*

The particular problem solved by the invention is the problem of training a large concatenated voice system. A large concatenated voice system with a large vocabulary is capable of speaking a number of different words. For each word in the vocabulary, the system has been trained so that a particular word has a corresponding phonetic sequence. Manual data entry is usually used to train these systems. This type of training technique is tedious, prone to errors, and has a tendency to be academic in entry style rather than capturing a true example

of how a word is pronounced or what a word, phrase, or sentence means or translates to. *Specification, p. 1, ll. 10-20.*

Claim 1 recites a method of training a computer system via human voice input from a human teacher. With reference to Figure 1, the computer system 10 has a text-to-speech engine 14 and a speech recognition engine 16.

With reference to Figure 3, the method comprises presenting a text spelling of an unknown word (block 60), and requesting to receive the human voice pronunciation of the unknown word using speech output (block 64). The request from the computer system 10 takes a form of an ongoing natural language dialog between the computer system 10 and the human teacher with the computer system 10 having a list of ways to ask questions with a variable for the questionable data (block 64).

The method further comprises receiving a human voice pronunciation of the unknown word from the human teacher (block 66), and determining a phonetic spelling of the unknown word with the speech recognition engine 16 based on the human voice pronunciation of the unknown word (block 68). The text spelling is associated with the phonetic spelling to allow the text-to-speech engine 14 to correctly pronounce the unknown word in the future when presented with the text spelling of the unknown word (block 70). *Specification, p. 2, ll. 4-13, ll. 18-20; p. 5, l. 4 - p. 6, l. 13; p.8, ll. 5-20.*

Figure 2 depicts an example of this ongoing natural language dialog between the computer system and the human teacher at 30. *Specification, p. 6, l. 14 - p. 7, l. 2.*

Claim 11 recites a computer readable storage medium having instructions stored thereon that direct a computer to perform a method of training a computer system via human voice input from a human teacher. The computer system has a text-to-speech engine and a

speech recognition engine. The medium further comprises instructions for presenting a text spelling of an unknown word, and requesting to receive the human voice pronunciation of the unknown word using speech output. The request from the computer system takes the form of an ongoing natural language dialog between the computer system and the human teacher with the computer system having a list of ways to ask questions with a variable for the questionable data. The medium further comprises instructions for receiving a human voice pronunciation of the unknown word from the human teacher, and instructions for determining a phonetic spelling of the unknown word with the speech recognition engine based on the human voice pronunciation of the unknown word. The text spelling is associated with the phonetic spelling to allow the text-to-speech engine to correctly pronounce the unknown word in the future when presented with the text spelling of the unknown word. *Specification, p. 3, ll. 15-27; p. 9, ll. 16-27.*

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1-2 and 11-12 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Baker et al. (U.S. Patent No. 6,092,044) in view of Beutnagel (U.S. Patent No. 6,078,885), and further in view of Junqua (U.S. Patent No. 6,598,018).

Claims 3 and 13 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Baker et al. in view of Beutnagel, further in view of Junqua, and yet further in view of Franceschi (U.S. Patent No. 6,321,196).

Claims 6-7 and 16-17 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Baker et al. in view of Beutnagel, further in view of Junqua, and yet further in view of Surace et al. (U.S. Patent No. 6,144,938).

VII. ARGUMENT

A. Claims 1-2 and 11-12 Are Patentable Under 35 U.S.C. § 103(a) Over U.S. Patent Nos. 6,092,044, 6,078,885, and 6,598,018

The invention comprehends, within a method of training a computer system via human voice input from a human teacher, requesting to receive the human voice pronunciation of the unknown word using speech output wherein the request takes a form of an ongoing natural language dialog between the computer system and the human teacher with the computer system having a list of ways to ask questions with a variable for the questionable data.

Baker describes a method of adding a word to a speech recognition vocabulary. The spelling and utterance of the word are received and a collection of possible phonetic pronunciations of the word are created. The collection is created by comparing the spelling to a rules list of letter strings with associated phonemes. Speech recognition could be used to find the pronunciation from the collection that best matches the utterance of the word. The Examiner acknowledges that Baker does not suggest the recited ongoing dialog feature.

Beutnagel describes verbal, fully automatic dictionary updates by end-users of speech synthesis and recognition systems. Beutnagel mentions that the system may ask the end-user to record another example of a target word, and then re-run recognition steps. The Examiner acknowledges that Baker in view of Beutnagel still does not suggest the claimed ongoing natural language dialog feature. In making the final rejection, the Examiner relies on Junqua as an additional secondary reference.

Appellant believes that there is no motivation to combine these references to achieve the claimed invention. The invention relates to a method of training a computer system via human voice input from a human teacher with the computer system including a speech recognition system. Junqua relates to an apparatus and method for using natural dialog to

control operation of an automobile system, such as a navigation system. Junqua is not in the field of Appellant's invention which is training a computer system via human voice input from a human teacher with the computer system including a speech recognition system. Although Junqua may use natural dialog to control operations of an automobile system, this technical area is different than the field of training a computer system with a speech recognition engine.

Further, Junqua would not logically have commended itself to an inventor's attention when considering the problem faced by Appellant. Appellant faced the problem of training a large concatenated voice system. The traditional use of manual data entry to train large concatenated voice systems has some shortcomings. Appellant was faced with the problem of developing a method of training a computer system that overcomes the shortcomings of traditional manual data entry methods. Junqua does not address this problem with training, and thus, in addition to Junqua being in a different field, Junqua does not address the problem addressed by Appellant. Put another way, Junqua is non-analogous art and there is no motivation to use Junqua in combination with the other cited references to achieve the claimed invention.

In the Advisory Action, the Examiner states that Junqua is analogous art since it is from a similar field of endeavor in speech recognition, or more specifically, spoken word recognition utilizing a natural language dialog.

The field of the invention is training, or more specifically, the field of training a computer system via human voice input from a human teacher with the computer system including a speech recognition system. This field of training is a specific field of endeavor, and Junqua is not within this field. The fact that Junqua mentions speech recognition and mentions natural language dialog does not place Junqua within Appellant's specific field of endeavor which is training a computer system via human voice input. Training is a very complex process and the fact that Junqua involves a recognition system does not instantly place

Junqua in the relevant field. Further, it is worth noting that the Examiner has not directed attention to any specific portions of Junqua directed toward training or solving problems with training.

The Examiner has only made a general statement that Junqua relates to speech recognition and natural language dialog and that this fact makes Junqua analogous art. Appellant disagrees and believes that Junqua is non-analogous art because Junqua is not in the field of training a computer system via human voice input, or pertinent to the problem of manual data entry training techniques in training large concatenated voice systems.

In addition to Junqua being non-analogous art, Junqua appears to also be deficient and not overcome the deficiencies of the other relied upon references. The Examiner makes reference to column 3, lines 31-41; column 4, lines 46-67; and column 5, lines 49-57. In column 3, Junqua does mention a sentence to synthesize including a fixed part and variable slots. Note, however, that a fixed part and variable slots would not result in the claimed ongoing natural language dialog. After all, the claims recite that the computer system has a list of ways to ask questions with a variable for the questionable data. Using a fixed part and variable slots as in Junqua does not result in the claimed feature because the fixed part in Junqua is fixed, as opposed to the claimed list of (that is, non-fixed) ways to ask questions.

At column 4, Junqua describes asking the user to provide information, but this does not suggest the specifically claimed approach of Appellant's invention. At column 5, Junqua does describe managing the turn-taking aspects of human-like back-and-forth dialog. Nevertheless, suggesting the turn-taking aspects of human-like back-and-forth dialog still does not suggest the particularly claimed approach using an ongoing natural language dialog between the computer system and the human teacher with the computer system having a list of ways to ask questions with a variable for the questionable data.

In the Advisory Action, the Examiner makes reference to these same portions of Junqua, and states that the question frame disclosed by Junqua utilizes a fixed part with the questionable data being variable. The claims recite “a list of ways to ask questions with a variable for the questionable data.” Junqua does not describe a list of ways to ask questions, but only describes a fixed part with a variable part. That is, in Junqua only the variable part changes where, according to the claimed invention, ongoing natural language dialog is achieved because the computer system has a list of (that is, non-fixed) ways to ask questions with the variable for the questionable data.

Claims 2 and 12 are dependent claims and are also believed to be patentable.

**B. Claims 3 and 13 Are Patentable Under 35 U.S.C. § 103(a)
Over U.S. Patent Nos. 6,092,044, 6,078,885, 6,598,018, and 6,321,196**

Claims 3 and 13 are dependent claims and are also believed to be patentable.

**C. Claims 6-7 and 16-17 Are Patentable Under
35 U.S.C. § 103(a) Over U.S. Patent Nos.
6,092,044, 6,078,885, 6,598,018, and 6,144,938**


The Examiner relies on secondary reference Surace as suggesting the incorporation of the information content level features recited by claims 6-7 and 16-17. Surace is believed to be non-analogous art. Appellant's invention relates to a method of training a computer system via human voice input from a human teacher with the computer including a speech recognition engine. On the other hand, Surace is in a different field and relates to voice user interfaces with personality. The particular problem solved by Appellant's invention is the problem of training a large concatenated voice system. Surace is not in the same field of the invention, and does not commend itself to one dealing with the problem of training a large concatenated voice system. Surace only describes voice user interfaces having personality.

Surace is believed to be non-analogous art; there is no motivation to combine these references to achieve the claimed invention.

The fee of \$500 under 37 C.F.R. § 41.20(b)(2) for filing a brief in support of an appeal is enclosed. The fee of \$120 under 37 C.F.R. § 1.17(a)(1) for a one month extension of time is also enclosed. Please charge any additional fee or credit any overpayment in connection with this filing to our Deposit Account No. 02-3978.

Respectfully submitted,

ELIOT M. CASE

By: 
Jeremy J. Curcuri
Registration No. 42,454
Attorney for Applicant

Date: January 26, 2006

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351

Enclosure - Appendices

VIII. CLAIMS APPENDIX

1. A method of training a computer system via human voice input from a human teacher, the computer system having a text to speech engine and a speech recognition engine, the method comprising:

presenting a text spelling of an unknown word;

requesting to receive the human voice pronunciation of the unknown word using speech output;

wherein the request from the computer system takes a form of an ongoing natural language dialog between the computer system and the human teacher with the computer system having a list of ways to ask questions with a variable for the questionable data;

receiving a human voice pronunciation of the unknown word from the human teacher;

determining a phonetic spelling of the unknown word with the speech recognition engine based on the human voice pronunciation of the unknown word; and

associating the text spelling with the phonetic spelling to allow the text to speech engine to correctly pronounce the unknown word in the future when presented with the text spelling of the unknown word.

2. The method of claim 1 wherein the phonetic spelling includes a sequence of phonemes.

3. The method of claim 1 wherein the phonetic spelling includes a sequence of known words.

6. The method of claim 1 further comprising:
establishing a plurality of request statements, each request statement having an information content level, the information content levels ranging from a low information

content level to a high information content level, the plurality of request statements being used by the computer system during the ongoing dialog.

7. The method of claim 6 wherein presenting, receiving, determining, and associating are repeated for a plurality of unknown words, and wherein the information content level for the request statements in the ongoing dialog progressively lessens as presenting, receiving, determining, and associating are repeated.

11. A computer readable storage medium having instructions stored thereon that direct a computer to perform a method of training a computer system via human voice input from a human teacher, the computer system having a text to speech engine and a speech recognition engine, the medium further comprising:

instructions for presenting a text spelling of an unknown word;

requesting to receive the human voice pronunciation of the unknown word suing speech output;

wherein the request from the computer system takes a form of an ongoing natural language dialog between the computer system and the human teacher with the computer system having a list of ways to ask questions with a variable for the questionable data;

instructions for receiving a human voice pronunciation of the unknown word from the human teacher;

instructions for determining a phonetic spelling of the unknown word with the speech recognition engine based on the human voice pronunciation of the unknown word; and

instructions for associating the text spelling with the phonetic spelling to allow the text to speech engine to correctly pronounce the unknown word in the future when presented with the text spelling of the unknown word.

12. The medium of claim 11 wherein the phonetic spelling includes a sequence of phonemes.

13. The medium of claim 11 wherein the phonetic spelling includes a sequence of known words.

16. The medium of claim 11 further comprising:
instructions for establishing a plurality of request statements, each request statement having an information content level, the information content levels ranging from a low information content level to a high information content level, the plurality of request statements being used by the computer system during the ongoing dialog.

17. The medium of claim 16 wherein presenting, receiving, determining, and associating are repeated for a plurality of unknown words, and wherein the information content level for the request statements in the ongoing dialog progressively lessens as presenting, receiving, determining, and associating are repeated.